Revised Guidance for Sample Sorting and Subsampling Protocols for EEM Benthic Invertebrate Community Surveys

1.0 INTRODUCTION 1

2.0 SAMPLE PROCESSING METHODS 1

3.0 LABORATORY SUBSAMPLING METHODS 9

3.1 When to subsample 9
3.2 Recommended subsampling approaches for EEM studies 9
3.3 Basic Laboratory Subsampling Methods 14
3.4 Calculation of subsampler correction factors, precision and accuracy 17

4.0 REFERENCES 19

ANNEX A 23

1.0 Introduction

The guidance documents prepared for both the pulp and paper and metal mining EEM programs provided guidance for laboratory sample processing, subsampling protocols and quality assurance. However, detailed protocols and acceptability criteria for subsampling were lacking. To identify if potential processing or subsampling issues exist within the EEM program, Cycle 2 EEM P&P interpretative reports were reviewed in detail. The information extracted from these reports included: the number of studies where laboratory subsampling was performed, the type of subsampling, if detailed protocols were provided, if density correction factors were reported and calculated appropriately, if subsampling error (precision and/or accuracy) was reported and calculated appropriately and finally if QA/QC criteria for sorting and subsampling were established and adhered to in these studies. The revised guidance provided in this document is a result of the inconsistency in the reporting of sorting and subsampling QA/QC information in these reports (see Cycle 2 review document for details). The intent of this revised guidance is not to re-invent established protocols, but to provide an additional level of detail to the current guidance documents, specifically pertaining to Sections 6.2.2 in the Pulp Paper Technical Guidance Document and 5.22 in the Metal Mining Guidance Document.

2.0 Sample Processing Methods

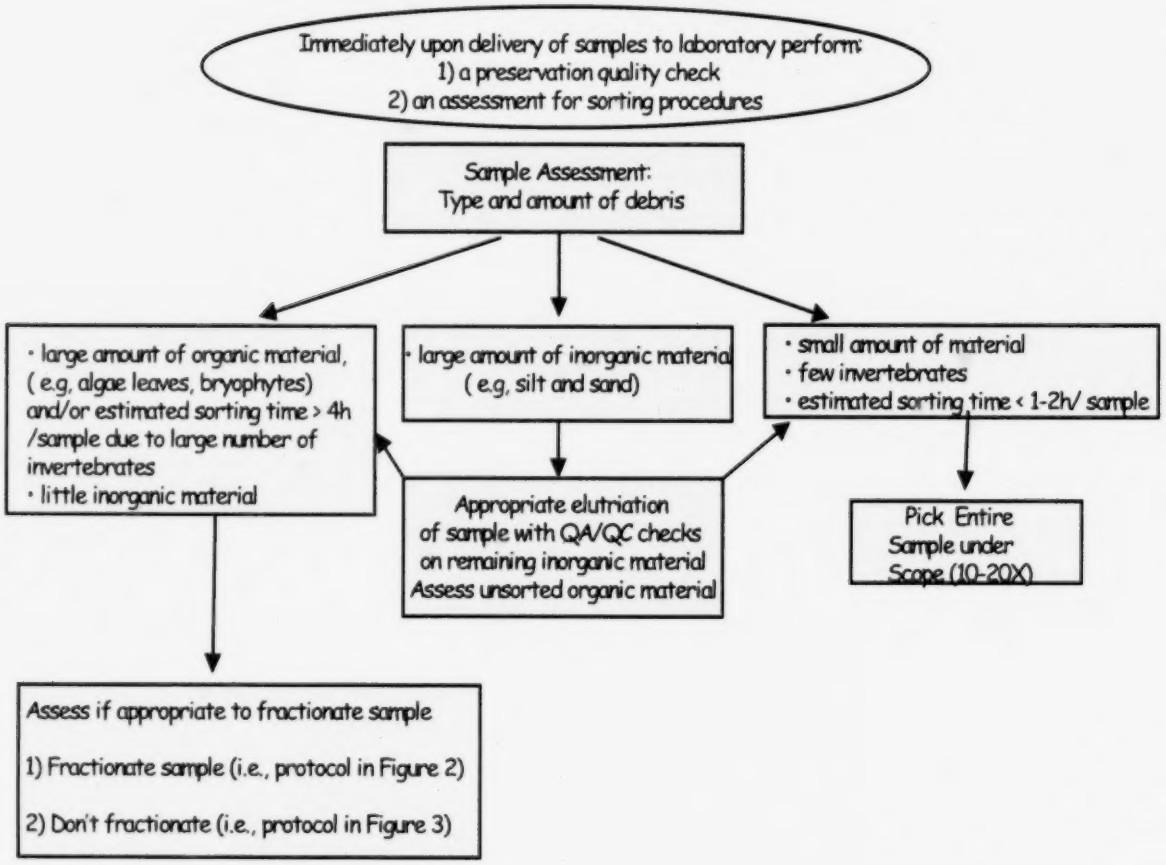
The sample processing steps recommended for all EEM benthic invertebrate sample processing are schematically represented in Figures 1-3. These are illustrative examples which would be applicable to most samples with decision points based on type and amount of organic and inorganic material. If sufficient numbers of organisms are encountered in a given fraction then a subsampling method can be considered (Figure 2 and 3, Section 3). Each laboratory should outline sample-specific sorting and subsampling procedures and provide full documentation of the methods used.

Preservation quality check and initial assessment of samples

Immediately upon return from the field, a quality check should be performed on the samples to ensure that the preservative has effectively penetrated the entire sample (Figure 1). This is especially important if there is a large amount of fine material in the samples. At this time, an initial assessment of the samples can be conducted to help determine the most efficient and cost effective method of sample processing. This assessment would include volume and type of debris material in the sample and if possible an assessment of the quantity of invertebrates in the samples. However, until sorting commences it is sometimes difficult to determine rough numbers of organisms in a sample. If sample size is small to moderate, sorting of the entire sample may be the simplest and most cost efficient method of sample processing. If large numbers of

organisms are present then subsampling may be warranted (see section 3.1 for determining when to subsample).

Figure 1: Summary of general sample processing protocols


Immediately upon delivery of samples to laboratory perform:
1) a preservation quality check
2) an assessment for sorting procedures
Sample Assessment:
Type and amount of debris
· large amount of organic material, (e.g, algae leaves, bryophytes) and/or estimated sorting time > 4h /sample due to large number of invertebrates
· little inorganic material
· large amount of inorganic material (e.g, silt and sand)
· small amount of material
· few invertebrates
· estimated sorting time < 1-2h/ sample
Appropriate elutriation of sample with QA/QC checks on remaining inorganic material Assess unsorted organic material
Pick Entire Sample under Scope (10-20X)
Assess if appropriate to fractionate sample
1) Fractionate sample (i.e., protocol in Figure 2)
2) Don't fractionate (i.e., protocol in Figure 3)

Washing, Elutriation and Storing of Samples

If samples have large amounts of inorganic matter (silt and sand), which was not removed in the field by bucket swirling and pouring the floating organic matter onto a sieve (Needham and Needham 1962), this procedure should be done in the laboratory as it can reduce sorting time by as much as 50% (Rosillon 1987, Ciboroski 1991). Furthermore, this elutriation process minimizes the adverse effects large amounts of inorganic matter may have on invertebrates (damaged specimens are more difficult to identify), or the subsampling process (easier distribution on sieves or trays). Before discarding any inorganic sample portions, a thorough visual inspection should be conducted to detect any cased or shelled invertebrates (e.g., rock encased caddisflies) that may be retained in the inorganic portion.

Prior to microscopic sorting of samples in the laboratory, sample preservative should be thoroughly rinsed from the organic matter. Sorting samples directly from a 10 % formalin solution is unnecessary and unwise from a health and safety perspective. The preserved sample can be placed into a sieve (freshwater = 500 µm) and the preservative and residual fine debris rinsed from the sample. Disposal of used preservatives must be performed in accordance to provincial hazardous waste regulations (e.g., BC Special Waste Regulations, Alberta Waste Control Regulation, Ontario Regulation # 347 etc.). While flushing the samples, care should be taken not to damage the organisms with too much water pressure. Samples can be sorted in water immediately, providing they are transferred back to preservative at the end of the workday (i.e., after 8 hours). Alternatively, samples may be transferred to 80% ethanol after the washing process, prior to sorting and taxonomic work provided that the sample has been thoroughly exposed to 10% formalin for a minimum of 72 hours. After sorting or identification, most freshwater macro-invertebrates should be stored in a solution of 70 to 80% ethanol, and 5% glycerin in vials or jars with air-tight lids. (Marine samples should be transferred to 70-80% ethanol within 3 months to protect the samples from breakage as the organisms become brittle.) If screw-cap lids are used, they should be sealed with wax film, and will require periodic checking (once or twice annually) to replace evaporative losses. Since preservatives can have a variety of effects on the lengths and weights of invertebrates (Howmineer 1972, Leuven et al. 1985), it is essential that the exact preservation protocol for all steps be drafted *a priori* and that it be consistently applied between all phases, areas, replicate stations and habitats.

Fractionation and Sorting of Samples

For samples, which have large amounts or pieces of organic matter, the samples can be divided in the laboratory into appropriate size fractions to expedite the sorting process (Figure 2). The most commonly used fractions are coarse (> 1.00 mm) and fine (500 µm - 1.00 mm), which correspond to the divisions used to define coarse and fine particulate organic matter (CPOM and FPOM, respectively). If there are very large pieces of organic material or large invertebrates it is sometimes beneficial to separate these from the rest of the sample with a 4.00-mm sieve. All fractions should then be sorted and if warranted, by large numbers of organisms, the fractions can be subsampled independently (see Section

3.0). Note that for samples with large amounts of clumping algae, fractionation may not be beneficial. Careful note taking is recommended for these more complex sorting procedures so that densities are calculated accurately. After the initial washing and fractionation of samples, the invertebrates should be sorted from the debris by trained technicians on a gridded tray or petri dish under a dissecting microscope at 10X to 20X magnifications.

Various additional techniques exist which are designed to speed up the manual sorting required for benthic invertebrate samples. These techniques include flotation of organisms in high-density solutions (e.g. magnesium sulphate, sodium chloride, calcium chloride, sugar, D-mannitol) or bubbling to trap animals in the surface film and to promote separation from denser sediment particles. Heavier organisms such as molluscs will not float and must be removed directly with forceps. Also, various stains such as rose bengal or phloxine B at a concentration of 100 mg/l can be used to make the organisms easier to see. A cautionary note is required, however, as the efficiency of many of these techniques differ between taxa (Rosillon 1987). If these techniques are used, then the QA/QC performed on these samples should be designed to detect any differential sorting efficiency of taxa caused by the staining procedures.

Figure 2: Sample processing with sieve fractionation

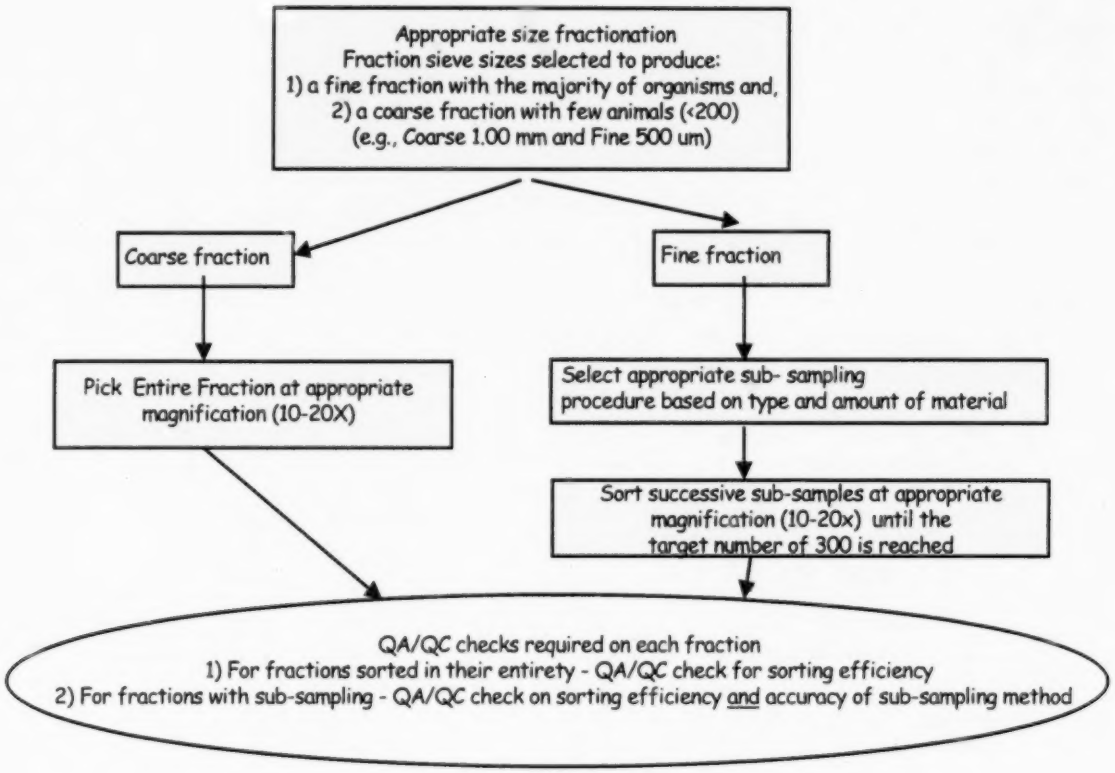
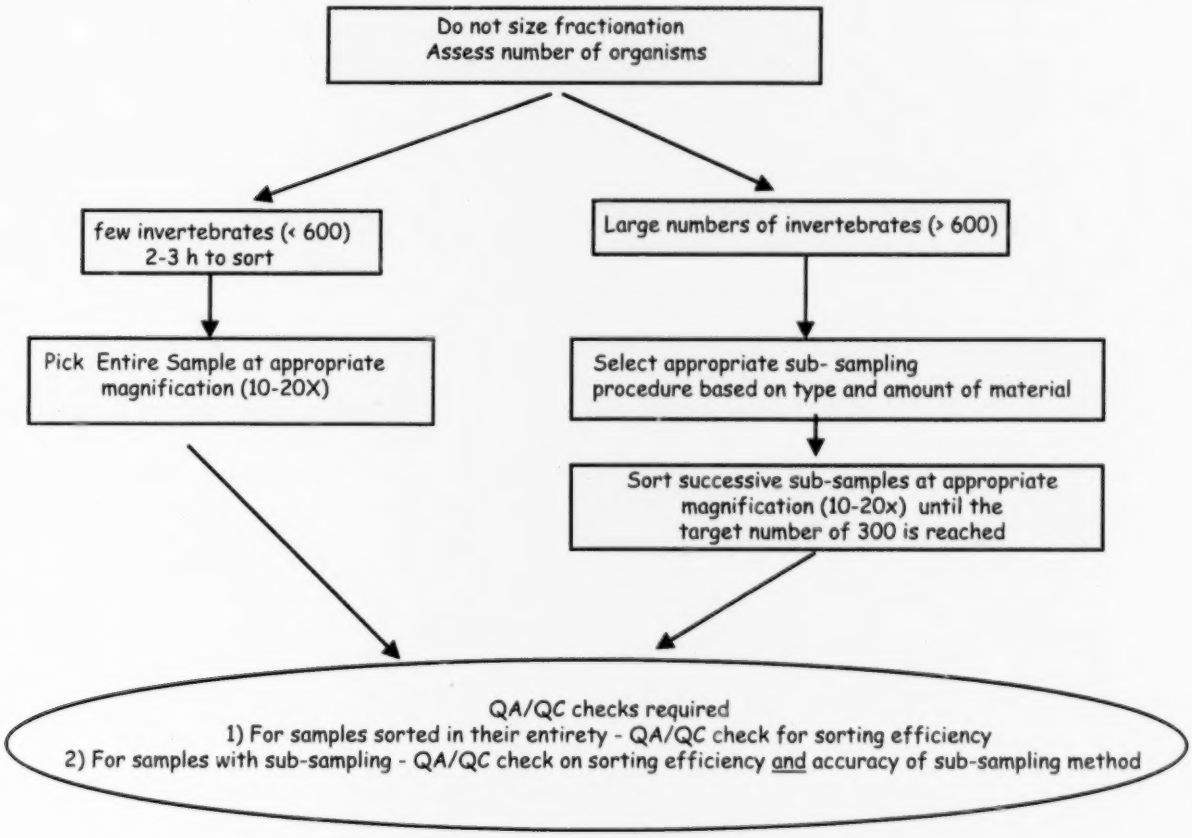

Appropriate size fractionation
Fraction sieve sizes selected to produce:
1) a fine fraction with the majority of organisms and,
2) a coarse fraction with few animals (<200)
(e.g., Coarse 1.00 mm and Fine 500 um)
Coarse fraction
Fine fraction
Pick Entire Fraction at appropriate magnification (10-20X)
Select appropriate sub- sampling procedure based on type and amount of material
Sort successive sub-samples at appropriate magnification (10-20x) until the target number of 300 is reached
QA/QC checks required on each fraction
1) For fractions sorted in their entirety - QA/QC check for sorting efficiency
2) For fractions with sub-sampling - QA/QC check on sorting efficiency and accuracy of sub-sampling method

Figure 3: Sample processing without sieve fractionation


Do not size fractionation
Assess number of organisms
few invertebrates (< 600)
2-3 h to sort
Large numbers of invertebrates (> 600)
Pick Entire Sample at appropriate magnification (10-20X)
Select appropriate sub- sampling procedure based on type and amount of material
Sort successive sub-samples at appropriate magnification (10-20x) until the target number of 300 is reached
QA/QC checks required
1) For samples sorted in their entirety - QA/QC check for sorting efficiency
2) For samples with sub-sampling - QA/QC check on sorting efficiency and accuracy of sub-sampling method

QA/QC for sample sorting

The basic processing of benthic invertebrate samples involves the time-consuming removal of organisms from large amounts of debris. Inevitably, processing errors occur during this sorting phase regardless of how diligent the processor and must be estimated (e.g. Kreis 1986, 1989). Thus, the first QA/QC component of benthic invertebrate sample processing is the requirement to assess this sorting efficiency (i.e., the proportion of total organisms extracted from the sample upon sorting). High sorting efficiencies will ensure that endpoint calculations are reasonably reliaole and without bias between samples. The recommendation for assessing sorting efficiency for the PP and MM EEM programs is that at least 10% of all samples from each study be resorted and any organisms found on the second sort be enumerated. The criteria for an acceptable sort is that > 90% of the total number of organisms are recovered during the initial sort. If > 10% of the total number is found during the resort, then all the samples within that particular group of samples requires resorting. The factors which should be considered when determining similar groups of samples include: 1) sampling area, 2) habitat class, and 3) individual sorters. The QA/QC guidelines apply independently to each sample fraction and group of samples sorted. A further criterion which would require a resort is if an entire group of benthic invertebrates were not extracted from the debris (e.g., ostracods were not recognized and sorted), even if the missed organism constituted < 10 % of the total numbers. Unsorted and sorted fractions are to be retained until taxonomy and sorting efficiency are confirmed and the data are reviewed by the regional EEM coordinator. The sorting efficiencies obtained on each set of samples should be calculated as below and reported for all studies. Note, that if the sorting efficiency is acceptable (>90%), the "resorted" organisms are left out of any further analysis as they are not part of the complete sorting process.

Determination of sorting efficiency

% Sorting Efficiency = [1- (# in QA/QC re-sort/(# sorted originally + # in QA/QC re-sort)]*100

Example:

- 37 invertebrates found in the QA/QC re-sort
- 393 invertebrates were recovered in the original sort process
- the sorting efficiency is calculated as 91.4% :

% Sorting Efficiency = [1- (37/(393 + 37))]*100

If the efficiency is less than 90% for any one group of samples, all samples within that group of samples should be completely re-sorted.

3.0 Laboratory Subsampling Methods

3.1 When to subsample

The processing of quantitative samples of benthic invertebrates is very laborious, largely because the amount of time required to extract large numbers of invertebrates from a large quantity of inorganic (sand and gravel) and organic (algae, leaves, bryophytes) material (Rosillon 1987, Ciborowski 1991). The two main factors which influence the sorting time of a given sample are the number of organisms and the amount of detrital material in the sample. In fact, these two factors explained close to 95% (84 and 10 % respectively) of the variation in sorting time in a series of samples examined by Ciborowski (1991). Samples with clumping Cladophora algae can result in a 1/3 increase in sorting time. A number of strategies have been developed over the years to reduce this processing time (Resh et al. 1985), however, these time saving methods must not be at the expense of data quality. Some processing time savers include elutriation, fractionation, flotation, and dyes and are discussed in the previous section (Section 2.0). These methods help to speed up the processing procedures due to large amount of organic or inorganic material. To reduce processing time due to large numbers of invertebrates in a given sample, a variety subsampling techniques have been developed. These techniques significantly reduce processing time and associated expense, while adequately estimating whole sample densities and taxonomic composition (Vinson and Hawkins 1996). The objective of any subsampling program should be to minimize effort (and hence cost) while yielding the maximum information with statistically reliable results (Wrona et al. 1982). Thus, the decision to use techniques such as fractionation are based on the amount and type of debris in the sample, while the decision to subsample is based on the number of organisms in the sample. These decision points are schematically represented in Figures 2 and 3.

3.2 Recommended subsampling approaches for EEM studies

The subsampling protocols in the following sections are recommended as a general approach to dealing with typical benthic samples from streams or lakes (Figures 2 and 3). However, the appropriateness of these general approaches will need to be assessed on an individual study, based on the type of material and numbers of organisms. The detailed reporting of subsampling accuracy and precision for all methods is essential to the QA/QC of EEM benthic invertebrate programs.

Sieve Fractionation and subsampling

For all samples for which subsampling is being considered, but especially for those which have large amounts or pieces of organic matter, it is strongly recommended that the samples be divided in the laboratory into appropriate size fractions to expedite the sorting

process (see Section 2.0). Sieving also produces a more homogeneous fraction, allowing for the efficient withdrawal of random subsamples (Anderson 1990) from the appropriate fraction (Taylor and Bailey 1997). The selection of sieve sizes for fractionation should include a consideration of separating the fine and coarse organic matter, but in addition, an attempt to capture the majority of the organisms in one fraction will enhance the subsampling reliability (Rossilon 1987, Meyer 1990). Thus, the selection of coarse sieve size can be related to the size of animals in the sample. Generally a sieve series of 1.00 mm and 500 µm (500 µm = recommended fine mesh and sieve size for EEM) is adequate but a larger coarse sieve of 2.00 or 4.00 mm may also be appropriate if there are many large invertebrates in the samples. The appropriate subsampling technique is then applied to the fine fraction as the majority of small invertebrates will reside on this sieve. However, for samples for which fractionation is not applicable (e.g., samples with clumping algae) appropriate subsampling methods (e.g., wet weight) are available which do not require fractionation (Section 2.3).

Minimum number of organisms

Inherent with subsampling is a potential reduction in the accuracy of endpoint estimates. The potential for error is greater with small subsamples or for less common taxa (Wrona et al. 1982, Meyer 1990). Many studies have reiterated the suggestion by Lund et al. (1958) that, based on the Poisson distribution, reasonable accuracy can be obtained when > 100 organisms are enumerated (Hickley 1975, Elliott 1977, Wrona et. al. 1982, Rossilion 1987, Klemm et al. 1990). Previous guidance indicated that subsampling should continue until a predefined variance level is obtained following the approach outlined by Wrona et al. (1982). For each study/ subsampler approach, an assessment would include a description of the trend in variance as number of animals sorted increased. Wrona et al. (1982) found that, for the Imhoff cone subsampler, as the total number of animals counted exceeded 50 and approached 100, improvement in the error term for a given unit of additional effort decreased. However, for counts of less common taxa this variance may be larger at a given number sorted (Wrona et al. 1982). For the EEM program a standardized minimum number of organisms to sort has not been explicitly stated. Thus, depending on the technique used or variance in the sample, minimum numbers could vary widely. With the national objective of comparing EEM studies across the country, and the recommendation of effect endpoints, including taxa richness, further standardization of subsampling protocols will be beneficial, including a minimum number of organisms.

Many studies have attempted to provide recommendations of a minimum number of organisms which provide reasonable subsampling accuracy and/or precision. Much of the literature deals with the fixed count methods as the number of organisms counted is obviously critical to this method. Minimum number initially recommended in the literature was set at 100 (Plafkin et al, 1989), while more recently, recommendations have ranged from 100-300 (Caton 1991, Hannaford and Resh 1995, Vinson and Hawkins 1996, Growes et al. 1997, Larsen and Herlihy 1998 Somers et al. 1998). In these recent assessments the endpoint of concern is usually taxa richness, as this metric is related to sampling (or subsampling) effort and the higher number (300) is generally recommended

(Barbour Gerritsen 1996, Vinson and Hawkins 1996, Larsen and Herlihy 1998, Somers et al. 1998). As taxa richness is one of the EEM benthic endpoints, , it is recommended that a *conservative approach* be taken to setting minimum numbers of organisms. Therefore, although the most important aspect of a subsampling program is the accuracy of the endpoint estimates, the recommendation that a minimum number of 300 organisms be removed from a sample in any subsampling programprovides additional standardization across all methods and studies for the EEM program.Note that if this minimum number is reached part way through a subsample sort, the subsample must be sorted in it's entirety so that the fraction sorted is quantitative. A minimum proportion of the sample to sort (e.g., 25%) could also be adopted, however, in very large samples this could still represent a significant workload (Ciborowski 1991) and is not recommended as it is the accuracy of the endpoint estimate that is the ultimate measure of any subsampling method.

Acceptable Error for Subsampling protocols

Regardless of the subsampling technique used, documentation of the accuracy of the estimate is essential to ensure that the data is comparable within and between studies. In fact, the main criteria to evaluate a subsampling technique is an evaluation of it's ability to accurately estimate the numbers and types of organisms in a sample. From the review of Cycle 2 Interpretative Reports, subsampling accuracy was generally not reported. Of those that reported some type of error, the majority reported the precision obtained when comparing two subsamples. For example;

1) a count in subsample A = 289
2) a count in subsample B = 316
3) the reported precision between theses two subsamples would be 8.5% (1-(289/316))x100.

If all the subsamples from this particular sample had similar precision, then the accuracy will also be close to 9%. However, without sorting the remainder of the sample, accuracy cannot be determined. The studies which did report the accuracy of the subsamples, did so as suggested in the guidance documents. For 10% of all samples several subsamples were sorted and ***then the remainder of the sample was sorted entirely***. Subsampling accuracy can then be calculated by comparing the estimates from the subsamples to the actual count. For example:

1) a count in subsample A = 289, representing 15% of the sample by volume, for an estimate of the total in the sample of 1927
2) a count in subsample B = 316, representing 15% of the sample by volume, for an estimate of the total in the sample of 2106
3) the count in the remainder of the sample = 1359, for a actual total of 1964
4) the reported precision would be the same as in the first example, 8.5 %
5) the reported accuracy would be -1.9% and +7.2% for sample A and B respectively.

Both precision and accuracy of the subsampling methods is information which is essential to ensure that the subsampling program is accurately estimating numbers of organisms in the sample. The overriding objective of subsampling is to reduce the substantial effort involved in processing benthic samples but not at the cost of data quality. For acceptable subsampling error criteria, the majority of Cycle 2 studies applied the 20% precision rule suggested by Elliott (1977). That is, if the precision between two subsamples was < 20% the error was acceptable (see example on precision above). Although this has been applied to the precision, it can equally be applied to the accuracy of the estimates. Although accuracy and precision obtainable depend on many factors, including the inherent variability of field samples (Norris et al. 1996), the standardization of EEM techniques should minimize this variability. Many researchers have suggested that is desirable to achieve estimates that are within 20% of the true count (Hickley 1975, Elliott 1977, Downing 1979, Wrona et al. 1982, Resh and McElvry 1993) and in fact, many of the subsampling techniques reviewed and recommended have demonstrated that at least this level of accuracy is attainable (Hickley 1975, Wrona et al. 1982, Rosillon 1987, Meyer 1990). Therefore, the criteria for an acceptable subsampling protocol is that the estimates of each group of samples should be within 20% of the true counts. The factors which should be considered when determining similar groups of samples include: 1) subsampling technique and 2) type of sample (i.e., type and amount of organic matter). As with sorting efficiency, the effects of subsampling on the accuracy of abundance estimates should be examined for a minimum of 10% the samples (or sample groups) in each EEM study and be appropriately reported (Table 1). If an acceptable error level is not demonstrated for a particular technique or a particular set of samples (e.g., with clumping debris, disallowing random mixing) then the technique should be modified to achieve this level of precision and accuracy or the sample should be sorted in it's entirety.

A note is required here regarding a reasonable level of effort to demonstrate accuracy. The critical point is that there is documentation available regarding the accuracy of the estimates, which was lacking in many Cycle 2 reports. The suggested guidance would essentially add one estimate of subsample accuracy to a standard EEM study (assuming similar sample type, 10% of 5 exposure + 5 reference samples = 1). There is no intention to have the documentation of subsampling accuracy become more onerous than sorting all samples. Furthermore, if the technique is one that is well established with a variety of sample types, published in the primary literature, then this exercise is simply a confirmation that the technique is being *applied* appropriately. In fact, a particular contractor may be able to demonstrate a technique's applicability for a set of samples with similar characteristics spanning more than one EEM study (e.g., samples processed within one season with the same operator(s)).

Table 1: Example of recommended reporting for subsampling error. Example is for a volume based method using the Imhoff cone (Wrona et al., 1982) where up to 10 subsamples are sorted and the remainder of the sample was sorted. Accuracy of each subsample, the minimum, maximum and mean subsampling accuracy as well as the range in precision between individual subsamples are all reported.

Subsample no	Number Inverts	Predicted no.	Pred - Expected	% Diff from
1	218	3815	101	2.7
2	220	3850	136	3.7
3	230	4025	311	8.4
4	221	3868	154	4.1
5	221	3868	154	4.1
6	201	3518	-197	-5.3
7	219	3833	119	3.2
8	205	3588	-127	-3.4
9	221	3868	154	4.1
10	210	3675	-39	-1.1
Total in remaining	1548			
Total in sample	3714			

		Mean Absolute subsampling error (%)	4.0
Range in Precision	0.5 - 9%	Min % error	1.1
		Max % error	8.4

3.3 Basic Laboratory Subsampling Methods

Once it has been established that subsampling will result in a significant timesaving in the processing of the samples with large numbers of invertebrates (usually greater than 500-600), an appropriate method can be selected. Various methods of laboratory subsampling are available and are outlined in Wrona et al. (1982), Sebastien et al. (1988), Marchant (1989), Plafkin et al. (1989), Klemm et al. (1990), Canton (1991) and Mason (1991). Regardless of the method selected, the recommended general protocols and QA/QC documentation (Section 2.1 and 2.2) should be followed as these aspects are critical to allow national comparisons of EEM studies.

There are two basic approaches for subsampling benthic invertebrate samples. The fixed count methods and fixed fraction methods (Barbour and Gerritsen 1996). The fixed count method is commonly used for enumerating plankton samples (Charles et al. 1991) and was adapted for sorting benthic samples by Hilsenoff (1987, 1988) and refined and modified for use with the US EPA Rapid Bioassessment Protocols (RBP) by Plafkin et al. (1989). Over the development of the method the total number of organisms to be counted has been debated (e.g., 100, 200, 300) (Plafkin et al. 1989, Caton 1991, Hannaford and Resh 1995, Growns et al. 1997, Doberstein et al. 2000, Carter and Resh 2001) but the basic premise is that a predetermined number of organisms are sorted and enumerated from any given sample. The original intent for using the fixed count methods with the RBP was to facilitate a rapid return of data for analysis and interpretation, allowing the identification of sites where potential further study would be indicated (Plafkin et al. 1989, Hannaford and Resh 1995). For these objectives, the fixed count methods are comparable between sites because a consisted unit of effort is applied. However, as Cortemanch (1996) points out, a fixed effort should not be confused with sample size. It is unknown whether the 100 organisms counted came from a sample of 500 or 5000 animals. Fixed count methods would be appropriate for proportional metrics (e.g., % EPT, Cortemanch 1996). However, as none of the benthic effect endpoints are proportional they would not be reliably estimated with fixed count subsampling methods. Therefore, although there are many literature reviews discussing the merits and applicability of fixed count methods, particularly for rapid bioassessment objectives, these methods would not meet the objectives of the EEM program where quantitative methods are required.

There are several fixed fraction subsampling methods, with the simple objective of dividing up a large sample, into several portions, each of which are a random representation of the entire sample. These smaller portions are processed more efficiently and cost-effectively while producing reliable estimates of the total number and types of organisms in the sample. The three most common methods divide up the samples based on either; 1) area, 2) weight or 3) volume of the sample. Which method is most appropriate depends on the type and amount of material in the sample. These methods are briefly reviewed below.

Area-based sieve splitting

Although there are several variations of area-based subsampling methods ranging from grided trays or frames to elaborate sieve splitting devices (Sodergren 1974, Hickley 1975, Klattenburg 1975, Rosillon 1987, Marchant 1989, Klemm 1990, Meyer 1990, Caton 1991, Mason 1991), the most commonly used in the EEM program have been based on the sieve diameter splitting method of Cuffney et al. (1993). The process is based on evenly distributing material on a tray or sieve and randomly separating a fraction (½, ¼ or 1/8) by planar area. Specifically, the sample is placed on a sieve marked with six equally spaced diameters. The sieve is immersed in water and agitated to produce an even distribution of material after which the water is drained. A dice is rolled to randomly select which of the six diameter markings to use for splitting the sample and then a ruler is used to divide the sieve into halves by aligning it with the marks and pressing the ruler edge into the sample. A scraper and wash bottle are used to help separate the sample into halves and remove the selected portions. Once the sample has been split, a coin toss is used to randomly select which half of the sample to sort. DO NOT discard the other half of the sample as it may be needed for QA/QC determination, or if the one half of the sample does not contain enough animals to fulfill the minimum 300 animal criteria it will need to be sorted. All unsorted and sorted fractions are to be retained until taxonomy and sorting efficiency are confirmed and the data are reviewed by the regional EEM coordinator. Very large samples may need to be split repeatedly to obtain suitable subsample fractions, but for each split, the corresponding half will need to be identified and saved for QA/QC purposes until all data is verified and accepted. An appropriate fraction(s) is sorted and the total number of organisms in the original sample is calculated by the inverse of the fraction (Section 3.4).

A new variation of this area-based subsampling method with a modified apparatus was recently developed by researchers at National Water Research Institute (NWRI) (Environment Canada 2002). Somewhat similar to the apparatus' and process described by Meyer (1990) and Mason (1991), it is capable of handling large samples typically collected in EEM programs and quantitatively divides samples into 4 quarters. The apparatus consists of an agitation rod and a 60-cm length of PVC pipe (8" dia.) with an attached (removable) sieve, pre-divided into quarters by an inset frame. The tube and attached sieve are immersed in standard 20L-bucket water to a depth of 40-50 cm. The sample is poured into the tube and agitated with the rod for a minimum of 30 seconds. The rod is removed, ensuring that no organisms adhere, and the entire apparatus is quickly lifted straight up from the water. As the water drains from the sieve, debris and invertebrates settle evenly over the sieve. The sieve is then detached from the tube and any material, such as filamentous algae, that straddles 2 quartiles are placed in the quarter to which the majority of the material resides. A prepared PVC template is placed over the sieve and one quarter of the sample is easily removed using either a spatula or wash bottle alone. Further quarters can similarly be removed from the sieve and stored appropriately for sorting or archiving. Tests for random splitting of samples have produced satisfactory results.

Volume based sample splitting

The Imhoff cone method detailed by Wrona et al. (1982), randomly distributes fine organic material and invertebrates in a 1 liter volume. The recommended protocol includes initially fractionating the sample into course and fine, producing a homogenous fine fraction with large numbers of organisms. This fraction is then placed in the cone apparatus and the volume brought to the 1 litre mark with water. A regulated air supply with an air stone siliconed into the base of the cone is used to gently agitate the sample for 2-5 min ensuring a random distribution of material. Note that inorganic material should have been removed previously in the elutriation step as it will congregate on the bottom of the cone reducing the effectiveness of the air stone. Approximately 50% of the sample can be removed from the cone with 55ml test tubes while continuously agitating the sample. The test-tube subsamples are sorted sequentially until the minimum 300 animals are found. As with the other subsampling methods, if this minimum number is reached part way through a subsample sort, the subsample must be sorted in it's entirety so that a known fraction is sorted. The portion of the sample remaining in the cone is retained along with contents of any unsorted test tubes as the unsorted fraction for potential further sorting for subsampling accuracy calculations. This method of subsampling is simple and effective for a wide range of benthic sample types. However, the technique may not be compatible with samples containing large amounts of filamentous algae.

Weight-based sample splitting

This method as documented by Sebastien et al. (1988) is a simple, reliable alternative for subsampling benthic invertebrate samples, especially those where macroinvertebrates are entangled in filamentous algae or bryophytes. The technique involves thoroughly mixing the sample in a 2-litre beaker to ensure random distribution and subsequently pouring and distributing the material evenly onto a pre-weighed sieve. The moist sample (and sieve) is weighed to the nearest 0.1g to obtain a total weight for the sample. Material is removed from the sieve at random in appropriate fractions (e.g., approximately 10% or 25% of material) and subsample weight recorded to determine the actual subsample fraction. Once the sample has been split, the first fraction to sort is selected randomly. Consecutive subsamples are sorted until the minimum of 300 animals is recovered and the particular fraction is sorted in it's entirety. The other fractions are retained for potential QA/QC determination.

3.4 Calculation of subsampler correction factors, precision and accuracy

Subsample correction factors and calculation of estimates of whole sample densities:

The correction factor and equation to convert subsample number to estimates for the entire sample are as follows:

1) Correction Factor (CF) = total sample area, volume or weight / subsample(s) area, volume or weight sorted to achieve the minimum 300 organism count

2) Whole density estimate = number of organisms in subsample(s) x CF

Examples:

Area-based methods:

Correction Factor = total area of sieve (415 cm^2)/ # of subsamples sorted x area of each subsampled (41.5 cm^2).

- 2 subsamples sorted to achieve 300 count.
- final count upon sorting the whole 2nd subsample in it's entirety is 347.
- Correction factor = 415 / 2 x 41.5 ml = 5

Estimate of total number of organisms in sample = 347 x 5 = 1735

Volume-based methods:

Correction Factor = total volume of sample in Imhoff cone (1000 ml)/ # tubes sorted x tube volume (55ml)

- 4 tubes sorted to achieve > 300 count.
- final count upon sorting the whole 4th tube in it's entirety is 365.
- Correction factor = 1000ml / 4 x 55ml = 4.55 .

Estimate of total number of organisms in sample = 365 x 4.55 = 1661

Weight based methods:

Correction Factor = total weight of sample 30.5 g (i.e., sample + sieve wgt – sieve wgt) / total weight of all subsamples sorted (11.7 g)

- 3 subsamples sorted to achieve > 300 count.
- final count upon sorting the whole 3th weight fraction in it's entirety is 419.
- Correction factor = 30.5 g / 11.7 g = 2.6.

Estimate of total number of organisms in sample = 419 x 2.6 = 1089

Determination of subsampling accuracy and precision:

The effects of subsampling on abundance estimates should be examined on a minimum of 10% of the samples. If the error exceeds 20% for any group of samples, all samples within that group of samples should be completely sorted to assure the subsampling process is not compromising data integrity. This requires that 10% of samples which have been subsampled are randomly selected and the remaining unsorted material is sorted in it's entirety. The estimates (calculated as above) are then compared to the actual counts from the sample and the accuracy of the estimates and the precision between subsamples can calculated as below:

Accuracy of the subsampling estimate

% Error in the estimate = [1-(estimated # in sample/actual # in sample)]x100

Example (repeated from Section 3.2)

1) a count in subsample A = 289, representing 15% of the sample by volume, for an estimate of the total in the sample of 1927
2) a count in subsample B = 316, representing 15% of the sample by volume, for an estimate of the total in the sample of 2106
3) the count in the remainder of the sample = 1359, for a actual total of 1964
4) the reported precision would be the same as in the first example, 8.5 % the reported accuracy would be -1.9% and +7.2% for sample A and B respectively.

Precision between subsamples

% Difference between two subsamples (A& B) = [1- (count in subsample A / count in subsample B)] x 100

Example (repeated from from Section 3.2)

1) a count in subsample A = 289
2) a count in subsample B = 316
3) the reported precision between theses two subsamples would be 8.5% (1-(289/316))x100.

4.0 References

Anderson, A.M. 1990. Selected methods for the monitoring of benthic invertebrates in Alberta Rivers. Environmental Quality Monitoring Branch, Environmental Assessment Division, Alberta Environment, Edmonton Alberta. 41p.

Barbour, M.T. and J. Gerritsen. 1996. Subsampling of benthic samples: a defense of the fixed count method. J. N. Am. Benthol. Soc. 15:386-391.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Rapid Bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates, and fish, 2nd Ed. EPA 841-B-99-002. US EPA, Office of Water, Washington, D.C.

Cao, Y., D. Williams and N. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? Limn. Oceanogr. 43:1403-1409.

Carter, J.L. and V.H. Resh. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. J.N. Am. Benthol. Soc. 20:658-682.

Caton, L.W. 1991. Improved subsampling methods for the EPA rapid bioassessment benthic protocols. Bull. N. Am. Benthol. Soc. 8:317-319.

Ciborowski, J.H. 1991. Estimating processing time of stream benthic samples. Hydrobiologia 222: 101-107.

Charles, D.F., S.S. Dixit, B.F. Cumming and J.P. Smol. 1991. Variability in diatom and chrysophyte assemblages and inferred pH: Paleolimnological studies of Big Moose Lake, New York. USA. J. Paleolimnol. 5:267-284.

Courtemanch, D.L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. J.N. Am. Benthol. Soc. 15: 381-385.

Cuffney, T.F., M.E. Gurtz and M.R. Meador. 1993. Guidelines for the processing and quality assurance of benthic invertebrate samples collected as part of the national water-quality assessment program. USGS OFR 93-406.

Doberstein, C.P., J.R. Karr and L.L. Conquest. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44:355-371.

Downing, J.A. 1979. Aggregation , transformation and the design of benthos sampling programs. J. Fish. Res. Board. Can. 36:1454-1463.

Eaton, L.E. and D.R. Lenat. 1991. Comparison of a rapid bioassessment method with North Carolina's qualitative macroinvertebrate collection method. J. N. Am. Benthol. Soc. 10:335-338.

Elliott, J.M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. 2nd. Ed. Freshw. Biol. Assoc. Sci. Publ. 25:160p.

Environment Canada. 2002. *(in prep).* A new cylinder subsampler for use with field or laboratory benthic macroinvertebrate samples. NWRI Contribution. No. XX-XXX., Environment Canada, NWRI, Saskatoon, SK.

Growns, J.E., B.C. Chessman, J.E. Jackson and D.G. Ross. 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. J. N. Am. Benthol. Soc. 16:682-692.

Hannaford, M.J. and V.H. Resh. 1995. Variability in macroinvertebrate rapid-bioassessment surveys and habitat assessments in a northern California stream. J. N. Am. Benthol. Soc.,14: 430-439.

Hickley, P. 1975. An apparatus for subdividing benthos samples. Oikos 26:92-96.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol. 20:31-39.

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. J. N. Am. Benthol. Soc. 7: 65-68.

Howmineer, R.P. 1972. Effects of preservatives on seights of some common macrobentic invertebrates. Trans. Am. Fish. Soc. 102: 743-746.

Klattenberg, R.P. 1975. A plexiglass subsampling box for large benthos samples. Progressive Fish Culturist 37:165.

Klemm, D.J., P.A. Lewis, F.Fulk, and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. US EPA/600/4-90/030

Kreis, R.G., Jr. 1986. Variability study Section 17. In: Charles, E.F. and R.R. Whitehead (eds.), Paleoecological investigation of recent lake acidification (PIRLA): methods and project description. Electric Power Research Institute Palo Alto California.

Kreis, R.G., Jr. 1989. Variability Study – Interim Results. Section 4. In: Charles, E.F. and R.R. Whitehead (eds.), Paleoecological investigation of recent lake acidification (PIRLA): 1983-1995: Interim Report. Electric Power Research Institute Palo Alto California.

Larsen, D.P. and A.T. Herlihy. 1998. The dilemma of sampling streams for macroinvertebrate richness. J. N. Am. Benthol. Soc. 17:359-366.

Lenat, D.R., 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertbrates. J. N. Am. Benthol. Soc. 7:222-233.

Leuven R.S.E.W., Brock, T.C.M., and H.A.M. van Druten. 1985. Effects of preservative on dry and ash-free dry weight biomass of some common aquatic macroinvertebrates. Hydrobiologia 127:151-159.

Lund, J.W.G., C. Kipling and E.S. Le Cren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia 11:143-170.

Marchant, R. 1989. A subsampler for samples of benthic invertebrates. Bull. Aust. Soc. Limnol. 12:49-52.

Mason, W.T. Jr. 1991. Sieve sample splitter for benthic invertebrates. J. Fresh.Ecol. 6:445-449.

Meyer, E. 1990. A simple subsampling device for macroinvertebrates with general remarks on the processing of stream benthos sampling. Arch. Hydrobiol. 117:309-318.

Needham, J.G. and P.R. Needham. 1962. A guide to the study of freshwater biology. Holden–Doug Inc., San Francisco. CA. 108p.

Norris, R.H., E.P. McElravy and V.H. Resh. 1996. The sampling problem. p. 184-208. In: Petts, G. and P. Calow (eds). River Biota Diversity and Dynamics. Blackwell Science Ltd. Oxford U.K.

Plafkin, J.L., M.R. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. U.S. EPA., EPA/444/4-89-001.

Resh, V.H., D.M. Rosenberg and J.W. Feminella. 1985. The processing of benthic samples: responses to the 1983 NABS questionaire. Bull. N. Amer. Benthol. Soc. 2: 5-11.

Resh, V.H. and E.P. McElvry. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. p 159-194 in: D.M. Rosenberg and V.H. Resh (eds.). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall. New York.

Rossilon, D. 1987. About the separation of benthos from stream samples. Arch. Hydrobol. 110: 469-475.

Sebastien, R.J., D.M. Rosenberg and A.P. Wiens. 1988. A method for subsampling unsorted benthic macroinvertebrates by weight. Hydrobiologia 157:69-75.

Sodergren, S. 1974. A simple subsample for stream-bottom fauna samples. Arch. Hydrobiol. 73: 549-551.

Somers, K.M., R.A. Reid, and S.M. David. 1998. Rapid Biological assessments: how many animals are enough? J. N. Am. Benthol. Soc. 17:348-358.

Taylor, B.R., and R.C. Bailey. 1997. Technical evaluation on methods for benthic invertebrate data analysis and interpretation, final report. AETE project 2.1.3. National Resources Canada, CANMET, Ottawa ON., 26pp.

Vinson, M.R. & C.P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. J. N. Am. Benthol. Soc. 15:392-399.

Waters, T.F. 1969. Subsampler for dividing large samples of Stream invertebrate drift. Limological Oceanogr. 14:813-815.

Wrona, F.J., Culp, J.M., & Davies, R.W. 1982. Macroinvertebrate subsampling: a simplified apparatus and approach. Can. J. Fish. Aquat. Sci. 39:1051-1054.

ANNEX A

Review of Cycle 2 Benthic Processing Protocols and QA/QC Issues

Nancy E. Glozier, J.M. Culp, Daryl Halliwell

Introduction

The technical guidance documents prepared for both the pulp and paper and metal mining EEM programs have provided guidance for laboratory sample processing, subsampling protocols and quality assurance. However, detailed protocols and acceptability criteria for subsampling have been lacking. Quality of laboratory benthic sorting can affect the accuracy of the benthic invertebrate community endpoints required to assess effects on fish habitat (i.e., total density, taxa richness, Simpson's Diversity and Bray-Curtis). If protocols within a single survey are consistent, minor errors in sorting efficiency are less problematic with regard to data interpretation of effects at a particular mill/mine. However, there may be a significant source of error in interpretation for individual mills/mines or for analyses performed at a larger scale (i.e., national or regional) if sample processing errors occur between samples, areas or mills.

To identify the scale of potential processing or subsampling issues the first step undertaken involved a detailed review of the methods reported in the Cycle 2 EEM P&P interpretative reports. The information extracted from these reports included: the number of studies where laboratory subsampling was performed, the type of subsampling, if detailed protocols were provided, if density correction factors were reported and calculated appropriately, if subsampling error (precision and/or accuracy) was reported and calculated appropriately and finally if QA/QC criteria for sorting and subsampling were established and adhered to in these studies. The intent of this review was not to re-invent established protocols but to review the level of detail in the current guidance documents and establish if the resulting interpretative reports provide the sufficient information to allow assessment of data quality. Areas where more specific guidance may be beneficial to the EEM program were identified, resulting in a revised draft of the guidance for benthic invertebrate sorting protocols, specifically pertaining to Sections 6.2.2 in the PP TGD and 5.22 in the MM GD. The following review covers three key issues reviewed from the Cycle 2 Interpretative Reports; sorting efficiency, subsampling methods, and subsampling precision / accuracy calculations.

Sorting Efficiency

The basic processing of benthic invertebrate samples involves the time-consuming removal of organisms from large amounts of debris. Inevitably, processing errors occur during this sorting phase regardless of how diligent the processor. The first QA/QC component of benthic invertebrate sample processing is the requirement to assess

sorting efficiency (i.e., the proportion of total organisms extracted from the sample upon sorting). High sorting efficiencies will ensure that endpoint calculations are reasonably reliable and without bias between samples. The recommendation for assessing sorting efficiency for the PP and MM EEM programs is that at least 10% of all samples from each study be resorted and any organisms found on the second sort be enumerated. The criteria for an acceptable sort is that ∃ 90% of the total number of organisms are recovered during the initial sort. In a review of the Cycle 2 Interpretative Reports 98% of all studies reviewed (87 mills) reported that this sorting QA/QC procedure was followed. However, only 58% of these mills reported the actual efficiency attained on the samples for a particular study. Of those reporting the efficiency, all attained the ∃ 90% target and most (75%) achieved ∃ 95% efficiency. Assuming, that the lack of reporting study-specific sorting efficiencies, does not represent efficiencies lower than the target, this aspect of benthic sorting QA/QC appears to be well understood and applied. However, sorting efficiencies should be reported for all studies in the standardized tables provided in the revised guidance.

Subsampling Methods

If the processing protocol involves subsampling of the sample, the second component of QA/QC involves an evaluation of subsampling error. As with basic sorting of benthic samples some level of error will result with any subsampling method. The error of concern for subsampling methods is how accurately the method estimates the total number of organisms in the entire sample. The similarity between two subsamples (i.e., the precision) is of less importance than the accuracy of the estimate. This said, if all subsamples were processed from a sample and the precision was high across all subsamples, the accuracy of each will also be high. However, if precision is low or variable, accuracy will vary with subsample. Thus, both of these measures of subsampling error should be evaluated with an appropriate QA/QC program.

To evaluate subsampling protocols in use for the EEM program, 83 mill studies were reviewed in detail. Close to 90% of all studies reviewed (83 mills) used some type of subsampler to assist in the processing of benthic invertebrate samples (Table 1). Although, subsampling methods vary, the simple objective is to divide up a large sample, into several portions, each of which are representative of the entire sample. These smaller portions are processed more efficiently and cost-effectively while producing reliable estimates of the total number of organisms in the sample. The three most common methods divide up the samples based on area, weight or volume of the sample and are appropriate for a range of sample types. The majority of Cycle 2 studies used an area based subsampling method. Although, there are a variety of area based techniques, all estimate the total number of organisms in a sample by the aerial proportion extracted off a planar surface. The other commonly used methods involved dividing the sample by weight or by volume. However, fully, 23 % of the studies did not indicate which method was used even though subsampling was used to assist in sample sorting. Several other techniques (Table 1, other category) were briefly described but their effectiveness is uncertain as they were not documented by literature evaluations. Finally, the fixed count

method, a subsampling method used widely in the US for field sampling programs, was not used in any of the EEM reports reviewed.

Table 1: Types of subsamplers used in the 83 Cycle 2 mill studies reviewed

Subsampler Method	**% Studies**	**Notes**
None	11	• entire sample sorted
Area	35	• % sample removed base on area • various methods based on Cuffney et al 1993
Weight	18	• % sample removed based on weight of debris • Sebastien et al 1988
Volume	10	• % of sample removed based on volume, animals randomly mixed in Imhoff cone • Wrona et al., 1982
Other	3	• Largely splitting the sample by other means, not documented by primary literature
Unknown/not specified	23	• methods not specified

Subsampling Precision and Accuracy

Regardless of the subsampling technique used, documentation of the accuracy of the estimate is essential to ensure that the data is comparable within and between studies. In fact, the main criteria to evaluate a subsampling technique is an evaluation of it's ability to accurately estimate the numbers and types of organisms in a sample. The reporting of subsampling error was reviewed from 57 mill studies from Cycle 2 P&P. Unfortunately over half of these studies (56%) did not report the error associated with the subsampling technique used. Of those that reported error, the majority (72%) reported the precision obtained when comparing two subsamples. For example;

1) a count in subsample A = 289
2) a count in subsample B = 316
3) the reported precision between theses two subsamples would be 8.5%
4) (1-(89/316))x100.

If all the subsamples from this particular sample had similar precision, then the accuracy will also be close to 9%. However, without sorting the remainder of the sample, accuracy cannot be determined. The studies which did report the accuracy of the subsamples, did

so as suggested in the guidance documents. For 10% of all samples several subsamples were sorted and then the remainder of the sample was sorted entirely. The subsampling error was calculated by comparing the estimates from the subsamples to the actual count. For example:

1) a count in subsample A = 289, representing 15% of the sample by volume, for an estimate of the total in the sample of 1927
2) a count in subsample B = 316, representing 15% of the sample by volume, for an estimate of the total in the sample of 2106
3) the count in the remainder of the sample = 1359, for a actual total of 1964
4) the reported precision would be the same as in the first example, 8.5 %
5) the reported accuracy would be -1.9% and +7.2% for sample A and B respectively.

This type of QA/QC information is essential to ensure that the subsampling program is accurately estimating numbers of organisms in the sample. A final note on the Cycle 2 Interpretative Reports is that, the correction factors used to estimate the total number of organisms in a sample were generally not reported. This is a simple calculation, especially for subsampling devices which simply divide the sample in half, but should, nevertheless, be reported.

In addition to reviewing the Cycle 2 interpretative reports, the guidance documents for both PP and MM were reviewed to assess whether further guidance may assist in obtaining more consistent reporting of QA/QC results. As there appears to be considerable variation in how error is calculated and reported a redrafting of the guidance pertaining to sections 6.2.2 in the PP and 5.22 in the MM guidance documents is underway. Included is a review and detailed description of subsampling techniques and suggested reporting procedures for QA/QC. An important addition to this guidance is a specific recommendation of an acceptable level of error for laboratory subsampling which was lacking in previous versions. Perhaps due to the lack of this criteria, none of the Cycle 2 studies which reported error (either precision or accuracy) outlined corrective action if error levels were unacceptable. The overriding objective of subsampling is to reduce the substantial effort involved in processing benthic samples but not at the cost of data quality. For the basic sorting efficiency, the acceptable error was set at 10%. For subsampling error the majority of Cycle 2 studies applied a 20% precision rule. That is, if the precision between two subsamples was < 20% the error was acceptable (see example on precision above). Although this level of precision was suggested in the guidance documents in terms of determining how many grab samples (i.e., field subsamples) should be taken at a given station, it was not *explicitly* suggested as an acceptable level for laboratory subsampling. An approach to setting these acceptability rules and a follow up course of action if the subsampling protocols fail these rules is included in the expanded guidance.